Parc d'Affaires Techn polis
3 avenue du Canada
Bâtiment Alpha, BP 810
Les Ulis – 91974 Courtaboeuf Cedex
France
Tel : 01 69 29 80 80
Fax : 01 69 29 80 79

Les Ulis, October 17th, 2003

To whom it may concern,

To our knowledge the present English translation of the French application N°0015838 filed on December 6th, 2000 by GenOdyssee (inventor: JL Escary) in French is complete and faithful.

Valérie MEYRIAL
Patent Counsel

PATENT

PATENT APPLICATION
No. 0015838
Filed in France
On December 6th, 2000

APPLICANT

GenOdyssee

TITLE

Process for determination of one or more functional polymorphism(s) in the nucleic sequence of a preselected functional "candidate" gene and its applications

PRIORITY CLAIMS

NO PRIORITY

The present invention concerns a process for determination of one or more functional polymorphism(s) in the nucleic sequence of a preselected "candidate" gene and its applications.

The study of polymorphisms in the human genome covers greater and greater importance especially for the search of the cause of certain diseases or particular sensitivities as well as for the search of medications allowing having an influence on them.

It is generally admitted that there is a genetic contribution and an environmental contribution to the appearance of common diseases in the human being and to the resistance of certain individuals to these same diseases. The predisposition and genetic resistance to the appearance of common diseases in the human being will be called hereafter "traits."

As for the genetic contribution to these diseases, two things are also commonly admitted by the one skilled in the art: on the one hand, the number of genes that participate for these traits in the human being is greater than one (polygenic origin of traits), and on the other hand, these traits are suspected to be attributable, in the majority, to variations in expression or function of the genes that are encoded on the human genome among the different individuals of the world population. These variations are also suspected by the one skilled in the art to be, in the majority, variations of a base pair or SNP (Single Nucleotide Polymorphism), which would represent on average a total of 0.1% of the sequence of the entire human genome i.e. nearly 3 million base pairs.

On the one hand, the characterization of the functional SNPs, which will reveal the presence of "candidate" genes alleles connected to a predisposition or to the development of common diseases in certain individuals, will make it possible to develop therapeutic molecules with the goal of correcting the observed effects of these alleles on the organism of carrier individuals and in particular, without being restricted to it, of correcting the impact of the

functional SNPs on the structure of the proteins encoded by the "candidate" genes in the patients.

Likewise, the functional SNPs that will demonstrate a relationship between the mutant alleles and a resistance of certain individuals to common diseases will make it possible to invent therapeutic molecules whose role will be to imitate the protective impact of these alleles on the organisms carrying these alleles, and in particular, to imitate the impact of these mutant alleles of functional SNPs on the structure of the corresponding carrier proteins.

These diagnostic/prognostic kits and these new therapeutic molecules will be the tools for prevention and treatment of common diseases.

The current efforts of post-genomic research relate to the search of functional SNPs that demonstrate the relationship between one or more mutant alleles and one of the two traits "sensitivity" or "resistance" to common diseases in the population. Thus, the search of new therapeutic targets on the genome such as described above is carried out by SNPs genotyping analyses in samples of persons preselected for one of the two traits, followed by statistical analyses of genetic associations between certain alleles encoded by these SNPs and the trait(s) of interest.

The individuals for whom the genotype must be determined are selected with the aid of precise phenotypic criteria such as for example medical, clinical, epidemiological, physiological or biological criteria, which measure the degree of sensitivity or resistance of these individuals to common diseases.

Therefore, up to now, the search of variations in the human nucleic sequences, especially those called SNPs (Single Nucleotide Polymorphisms) that is, concerning one nucleotide, has been carried out either systematically (sequencing of the human genome) or by proceeding with sequencing of the genome of individuals selected, for example, because of a particular sensitivity or resistance that they present.

The method used consisted of discovering a direct relationship between a mutant allele encoded by a functional or nonfunctional SNP and one of the two traits of common diseases.

This is done in four steps:

- in step 1 one proceeds with the identification of SNPs in a sample of patients and/or in a sample of resistant people and, always, in a sample of individuals known as controls (individuals presenting normal phenotypic data regarding the trait(s) studied). Furthermore, the SNPs are searched either on the genome in order to determine an association or genetic linkage between one or more regions of the genome and the trait(s) studied ("Genomescan" approach),
- step 2 consists of genotyping alleles encoded by the SNPs identified in the first step in a sample of patients and/or resistant people, and always, also, in a sample of controls, followed by statistical analysis of the associations or genetic linkages between one or more genotyped allele(s) and the trait(s) studied.
- in step 3 the genotyping data are analyzed as follows: statistical calculations that make it possible to estimate the degree of reliability of a genetic association noted by a higher frequency of one or more allele(s) in the individuals selected for one or the other of the traits than in the control individuals. The genetic associations between one or more functional SNP(s) and one or the other of the traits, which are confirmed by the statistical calculation, reveal a relationship between the variability of expression or function of the gene(s) and protein(s) carrying the SNP(s) and the trait studied. This information makes it possible to give the status of therapeutic targets to mutant alleles of the genes concerned. Recent deciphering of the sequence of the human genome and the sequencing of numerous new genes on the genome make it possible to imagine to identify, in the near future, numerous new therapeutic targets according to this method for the prevention and treatment of common diseases.
- step 4 consists of confirming the status of therapeutic targets for certain alleles encoded by functional SNPs identified as genetically associated with the trait of interest. This is done by developing biological tests that make it possible to establish, by a modeling method, the relationship between the

allele and the trait. For example, it is shown that the mutant allele encoded by one SNP found in the promoter region of a "candidate" gene has an effect on the expression of the gene, or also that the mutant allele encoded by a functional SNP found in the coding sequence of a "candidate" gene has an effect on the structure of the protein encoded by the gene, and even more on the structure of the active domains of this structure, showing a clear effect of the mutant allele on the activity of said protein and therefore of the gene. The biological information created is indispensable to be able to make a functional link between the genetic study of the trait and, without being restricted to these specific data, the medical, clinical, physiological or biological data collected to select the patients or resistant people according to the trait studied. From this functional link established between certain alleles and the trait studied, and the characterization of the biological impact of the allele concerned on the expression or the function of the gene or protein studied, diagnostic/prognostic kits and/or new therapeutic molecule(s) can be developed.

In addition, the search of patients for whom a genetic character must be determined requires long, expensive and often difficult operations, aiming to form phenotypic groups of interest from which the DNA sequences must be studied. This is especially due to the fact that it is necessary, prior to starting up the study, to search and find a representative number of persons manifesting a common phenotypic character.

It would be desirable to provide a method making it possible to discover the existence of polymorphisms in the human genome with good certainty.

In addition, systematic sequencing leads to a significant loss of energy since it also amounts to working on sequences without value, especially therapeutic value.

Now, the applicant has identified a new method enabling to find polymorphisms and especially genomic defects that especially present the following advantages:

Without resorting to genotyping studies of persons presenting a particular phenotype and to association studies or studies of genetic linkage between SNP markers and the phenotype(s) studied, which follow, the process makes it possible to form a databank of genetic variants responsible for functional changes in the expression or in the activity of the genes on the genome, and therefore, diagnostic/prognostic and potential therapeutic targets on the genome for the prevention and treatment of common diseases. Indeed, it is recognized that the impact of the gene pool of a person on his (her) sensitivity or resistance to the appearance and to the development of the diseases is due to mutations that change the normal expression and/or the normal activity of one or more of his (her) genes. The functional SNPs are counted among these mutations. Among them, one or all will, therefore, form targets for the development of diagnosis/prognosis and therapeutic kits for the prevention and treatment of said diseases.

Furthermore, the process is more reliable for discovering prognostic/diagnostic and therapeutic targets on the genome by comparison with statistical studies of associations or genetic linkages carried out with the help of genotyping studies on persons sensitive or resistant to the diseases and control persons. In fact, although measured, the risk is real of discovering an association or a genetic linkage between one or more SNPs and the appearance and/or development of one or more disease(s) while this association or genetic linkage is false in reality (this type of association or genetic linkage is called a false positive association or linkage) and cannot be avoided owing to the very statistical nature of the methods of calculation.

Because of this, the present process describes the development of concrete biological tests demonstrating the real functional role of certain alleles encoded by functional SNPs on the expression or activity of genes and constitutes a more reliable discovery of potential diagnostic/prognostic and therapeutic targets on the genome.

The process according to the invention also makes it possible to economize with any preselection of persons for a particular phenotypic trait,

defined here as a particular sensitivity or resistance to diseases, to discover functional SNPs forming potential diagnostic/prognostic and therapeutic targets on the genome. The process of the invention therefore makes it possible to save time, money and energy in the discovery of these potential targets for the development of diagnostic/prognostic kits and therapeutic molecules for the prevention and treatment of diseases.

The process according to the invention is based, in contrast to prior art, on the identification of functional SNPs in "candidate" genes, in a random population not selected on (and without being restricted to these particular criteria and data) medical, clinical, epidemiological, physiological or biological criteria and data. In other words, the process according to the invention relates to a method that makes it possible to discover functional SNPs in "candidate" genes in a random population, enabling the identification of mutant alleles forming potential therapeutic targets or so-called "candidate" therapeutic targets for the diagnosis/prognosis or treatment of common diseases, without resorting to the analysis of samples from preselected patients or resistant individuals. This random population takes into account a large number of different human ethnic groups.

The process is carried out in simply two major steps: the identification of the genotyping of functional SNPs in a random sample of the population composed of individuals recruited at random in the population, and the biological validation of the impact of the mutant allele encoded by each of the functional SNPs on the expression or function of the "candidate" genes or proteins encoded by these genes.

The identification of a strong biological effect of these alleles on the expression or the function of the "candidate" genes or the proteins encoded by these genes makes it possible to attribute, with the help of data available in the prior art concerning the functional "candidate" genes, the status of potential or "candidate" therapeutic targets to mutant alleles demonstrating a strong biological effect, this status being attributed for therapeutic fields (common diseases) for which, according to the prior art, the "candidate" genes are

suspected of playing a part.

Once the SNPs are detected, the identification of the allele(s) genetically associated with the trait(s) of interest and, therefore, the identification of new therapeutic targets related to common diseases can be carried out.

As common diseases are by definition diseases that concern a large number of individuals, a sample of individuals taken at random in the population therefore contains a reasonable number of patients and resistant persons not identified as such. Thus, functional SNPs can be discovered that are associated with one or the other of the traits of common diseases and, therefore, making it possible to identify therapeutic targets related to these diseases by directly analyzing such a population of individuals known as random population. The genotyping of these same individuals for the functional SNPs so identified makes it possible to estimate the allele frequency of these SNPs in the different human ethnic groups represented in the random population, which also makes it possible to predict the impact of the identification for the diagnosis/prognosis or treatment of these different ethnic groups.

That is why the present document claims as an object a process for determination of one or more functional polymorphisms in the nucleic sequence of a preselected "candidate" gene in which:

a) the genomic nucleic acid fragment of the "candidate" gene is isolated from a significant number of individuals chosen randomly in the population,

b) a comparative analysis of the nucleic sequence of the individuals studied is conducted,

c) the identical nucleic sequences are classified into homogeneous groups, and

d) the polymorphism of the nucleic sequence of each group is identified by comparison with the nucleic sequence of the reference "candidate" gene.

Thus, instead of proceeding with a systematic work as in the prior

art and from specific individuals (patients or resistant persons) to get their genes and to study them, the process of the present invention starts only with genes known in the prior art as fulfilling particular functions in a pathology or in a particular biological process, and the genes studied are from a random population sample, i.e. one that is not chosen because it presents the character one is trying to study.

In the present invention and in what follows, the "candidate" gene is designated as a gene where the following is known:

- all or part of the regulatory and coding nucleotide sequence and/or the sequence of the protein encoded by this gene, and
- the knowledge of any medical, clinical, epidemiological, physiological or biological data relative to said nucleotide sequence or to said protein and which makes it possible to reveal to the experimenter, a potential or assumed role of the expression of these genes or of the protein(s) encoded by these genes, if it or they exist, or also the activity of the protein(s) encoded by these genes, if it or they exist, in the appearance of common diseases or, on the contrary, in a particular resistance to these diseases in the human population.

"Functional candidate gene" is understood to be a "candidate" gene for which the function can be determined. "Functionality" is understood to be the modification of the biological activity of a biological molecule, this modification consisting in an increase, decrease or suppression of said biological activity. The biological activity can especially be linked to the affinity or absence of affinity of the biological molecule with a receptor.

"Reference wild-type sequences" are defined as regulatory and coding nucleotide sequences of the "candidate" gene as defined above, and which are known entirely or in part in the prior art and which act as templates for the experimenter for the design of fragments of the "candidate" gene and the PCR amplification (Polymerase Chain Reaction) of these fragments from the genomic DNA of the individuals of the random population to carry out the identification of the functional SNPs in these individuals. Also included as a

reference wild-type sequence is the sequence of the protein encoded by the reference wild-type sequence of the "candidate" gene such as defined above and which is either known in the prior art or determined by the experimenter from the reference wild-type coding sequence of the "candidate" gene such as defined above and known in the prior art.

It is also acknowledged that in the case where the reference wild-type sequence of the "candidate" gene is not entirely known from the prior art, the one skilled in the art, with his own technological resources including, for example, cloning and sequencing of all of the regulatory and coding sequences of the "candidate" gene, from complete or partial sequencing of a genomic clone containing all or part of the sequence of the "candidate" gene, can determine the missing part and integrate it with the identification of the functional SNPs in the "candidate" gene within the random population.

"SNP" designates any natural variation of a base pair identified in a "candidate" gene in the genome of one or more individuals within the random population. Are preferably designated the SNPs identified only in the regulatory sequences containing, for example, the promoter, the potential "enhancer" sequence(s) and the splicing sites of the introns of the "candidate" gene or also the coding sequence (the exons) of the "candidate" gene. Each SNP reflects the presence of two different bases in the same position in the nucleotide sequence of the "candidate" gene, demonstrating the presence of two different alleles of the "candidate" gene in the genome of the individual or individuals in which the SNP has been identified in the random population.

"Functional" SNP is any natural sequence variation of a base pair in the regulatory sequences of a "candidate" gene or, if it exists, in the coding part of the gene sequence that codes for the signal peptide of the protein(s) encoded by the "candidate" gene, which is identified in the genome of one or more individuals of the random population and which reveals a variability in the expression of the "candidate" gene (level of transcription and translation) or of the protein(s) encoded by the gene if it or they exist (post-translational modifications such as glycosylation for example) in the random population.

"Functional" SNP is also any natural variation of a base pair situated in the coding sequence of a "candidate" gene and identified in the genome of one or more individuals of the random population which reveals either a stopping of translation (introduction of a STOP codon) or a modification in the nature of an amino acid in the protein(s) encoded by this gene if it or they exist and which changes the activity of said protein(s), revealing a variability in activity (also called functionality) of the protein(s) encoded by the "candidate" gene in the random population. This latter type of "functional" SNP is distinguished from the SNP known as "coding" which is formed by any natural variation of a base pair identified in the coding sequence of a "candidate" gene in the genome of one or more individuals of a random population and which causes a change in the nature of an amino acid in the protein(s) encoded by this gene if it or they exist and which does not change the activity of said protein(s). The functional and coding SNPs are distinguished from the SNPs known as "silent" also identified in the coding sequences of the "candidate" genes in the random population but which do not change the nature of the amino acids in the proteins encoded by these "candidate" genes.

The "candidate" functional gene can be preselected by carrying out a search in the literature (NCBI, Entrez or Medline, for example) and the databases (PubMed or OMIM, for example). The extrapolation of data obtained in models other than the human model (murine, yeast, etc.) is possible but necessitates the characterization of the human genes/proteins involved in the processes described in these models (for example: by sequence homology, by reconstruction of signaling pathways or metabolic pathways).

By definition, "mutant" or "mutated" sequence is any regulatory or coding nucleotide sequence of the "candidate" gene corresponding to a new allele of the gene revealed by the identification of a SNP in these sequences and that is unknown in the prior art. Likewise, mutant or mutated sequence is any new sequence of the protein encoded by the "candidate" gene that is revealed by the identification of a coding SNP in the coding sequence of the "candidate" gene and that is the expression of a new allele of the gene coded

by the coding SNP and that is unknown in the prior art.

"Common" disease is any disease in the human population for which it is thought that more than one gene is involved in its appearance in patients and/or in a particular resistance to the development of this disease in certain individuals of the population. Such diseases are also called, for the same reasons, polygenic diseases. These are, among others, the cancers, cardiovascular diseases, any disease forming a risk factor for the cardiovascular diseases such as, for example, diabetes type 1 and 2, hypertension, hypercholesterolemia, metabolic diseases such as obesity, also the autoimmune diseases, infectious diseases, diseases of the central nervous system such as for example Alzheimer's disease or schizophrenia or also depression, also the rejection of tissue(s) or organ(s) graft, anemia, allergy or also asthma.

The "candidate" functional gene is first chosen according to the prior art that allows determination of its potential role in the appearance of common diseases in the human population or in a particular resistance to these diseases by certain individuals in this population.

Next, the nucleic sequence of the "candidate" gene is isolated from a random population of a significant number of individuals.

"Random population" is defined as any human population where the individuals have been recruited at random and without particular phenotypic criteria including, for example, the collection of medical, clinical, epidemiological, physiological or biological data.

In a following step, the genes prepared as described above are subjected to a qualitative and quantitative analysis such as chromatography to detect a genotype and/or sequence difference between the different molecules of DNA studied.

Next, the identical nucleic sequences are classified in homogeneous groups (by alleles).

Then, one proceeds with the sequencing of the nucleic sequences of each group according to methods well known in the state of the art.

Then, if desired, one proceeds with the genotyping of the nucleic sequences of each group.

The process of the invention is illustrated by the interferon α 2 case in which a functional SNP has been identified in the coding part of the gene and which reveals a strong change in the structure of the binding site of interferon α 2 to its receptor.

The prior art has already revealed the essential role of this site in the function of interferon α 2 and makes it possible to predict a strong role of the mutant allele analyzed here in the function of interferon α 2. The prior art also shows the important role of this gene as immunomodulator and essential agent for the response of the organism to infection by a large number of infectious agents (viruses, bacteria, fungi and parasites).

Interferon α 2 is currently used as therapeutic agent to treat various types of cancers as well as to fight infection by the Hepatitis B and C viruses and the AIDS virus. These data make it possible to give a probable status of potential or candidate therapeutic target to the natural mutant allele identified in the random population and causing a major modification in the structure of the active site of interferon α 2.

The present invention especially has as an object a process for determination described above, in which the gene is preselected by carrying out a search in the literature or databases such as NCBI, Entrez or Medline for example, and PubMed or OMIM for example, respectively. The extrapolation of data obtained in models other than the human model (murine, yeast etc.) is possible but necessitates the characterization of the human genes/proteins involved in the processes described in these models (for example: by sequence homology, by reconstruction of signaling pathways or metabolic pathways).

The functional gene is also preselected by carrying out a search in the literature or in databases such that the following could be described in them: for example, the reference wild-type sequence of the gene and of the protein(s) encoded by this gene in the human being and/or in any species of the animal kingdom, the structure of the reference wild-type protein(s) in the human being

and/or in any species of the animal kingdom, one or more studies of the structure of the reference wild-type protein(s) encoded by the candidate gene such as crystallography studies, one or more comparison studies of the sequence of the reference wild-type gene in the animal kingdom, one or more experiments of site-directed mutagenesis on the reference wild-type sequence of the candidate gene showing the role of certain amino acids in the function of the protein(s) encoded by the candidate gene, activity tests performed *in vivo* in animals or *in vitro* with human or any other animal's cells such as for example tests for cellular proliferation, differentiation, or showing the involvement of the reference wild-type gene or protein in the activation or repression of a metabolic pathway, in particular the regulation of the activity of protein kinases and the nuclear expression of particular genes, animal models showing the role of the gene or of the protein(s) encoded by the "candidate" gene in the appearance of a particular pathology (for example transgenic mice), epidemiological, medical or clinical data showing an involvement of the gene or the protein(s) encoded by this gene in the appearance of or resistance to a common disease in the human population.

Thus, the "candidate" gene is chosen according to the prior art. It makes it possible to determine its potential role in the appearance of common diseases in the human population or in a particular resistance to these diseases by certain individuals in this population.

Any gene of the human genome known in the prior art, the understanding of which, either published in the literature or not, suggests or shows to the one skilled in the art to have a potential role through either the expression of this gene (transcription or translation level) or of the protein(s) encoded by this gene if it or they exist (post-translational modifications), or also the activity of the protein(s) encoded by this gene if it or they exist, in the appearance of common diseases or on the contrary, in a particular resistance to these diseases in the human being is considered as a "candidate" gene accessible to the one skilled in the art through different sources. These gene sequences described in the literature are called "reference wild-type

sequences."

Among the data of the prior art that may be used for the identification and characterization of functional SNPs in the "candidate" genes in the random population, particular attention is given to the knowledge of regulatory sequences of the "candidate" genes and, if they exist, sequences that, in the coding sequences, code for signal peptides of the proteins encoded by these genes that are responsible for the expression of these genes or protein(s) encoded by these genes, and to the knowledge of the three-dimensional structure of the reference wild-type proteins encoded by the reference wild-type coding sequences of the "candidate" genes, as well as to the knowledge of the amino acids that have been identified within these structures as taking part in the activity of said reference wild-type proteins.

A process in which the "candidate" gene is relevant in a particular pathology is preferred.

The "candidate" gene can especially be any gene likely involved in biological processes or common diseases, or in a particular resistance to these diseases in the human being, very particularly the human interferon α 2 gene.

On the other hand, the individuals can be selected by ethnic groups as will be seen hereafter in the experimental part, and for each of these groups a "significant number of individuals" per ethnic group can be taken, thus forming the random population, for example greater than 5, especially greater than 10, preferably greater than 20 and very particularly greater than 100.

"Significant number of individuals" is understood to be a number of individuals and therefore of genes studied for example, greater than 100, especially greater than 150, preferably greater than 200 and very particularly between 250 and 400.

Under preferred conditions of carrying out the above process, the nucleic sequence of the "candidate" gene of a significant number of individuals chosen randomly in the population is isolated by a PCR reaction. The Polymerase Chain Reaction is well known to the one skilled in the art.

The isolation of genomic DNAs can also be carried out by

methods well known in the state of the art.

Under preferred conditions of carrying out the above-described process, the specific DNA fragments corresponding to the predetermined fragments of regulatory and coding sequences of the "candidate" genes of individuals of the random population are amplified by Polymerase Chain Reaction (PCR) using appropriate oligonucleotide primers. Softwares such as Primer3® can be used to choose several pairs of primers to amplify the chosen regions by PCR (for example total or partial binding sequences for transcription factors in the promoters, total or partial splicing sequences of introns, total or partial sequences of exons).

Especially in the case of the interferon α, the Polymerase Chain Reaction is carried out from primers corresponding to the sequences ID SEQ No. 1 and ID SEQ No.2.

While the comparative analysis of the nucleic sequence of the individuals studied can be carried out by any technique known to the one skilled in the art, denaturing high performance liquid chromatography (DHPLC: "Denaturing-High Performance Liquid Chromatography") is particularly preferred.

Under preferred conditions, the detection of the SNPs is carried out by DHPLC analysis. This methodology makes use of the fact that double-stranded homo- and hetero-duplex species are differently retained on a column under conditions of partial thermal denaturation.

Indeed, DHPLC presents the advantages of detecting SNPs with a greater effectiveness (97%) by comparison with sequencing (85 to 90%).

Such a process which involves the use of a multiplexing method of samples is described in FR-A-2 793 262 (Application No. 99 5651 of May 4, 1999).

Briefly, the DNA fragments amplified from the genomic DNA of heterozygous or homozygous individuals are separated under partially denaturing conditions by HPLC.

Preferably, the amplification products corresponding to several

individuals, preferably between 3 and 50 individuals, particularly between 3 and 5 individuals, and very particularly 3 individuals, are mixed before proceeding with the denaturation and DHPLC analysis.

Other preferential conditions for carrying out the DHLPC and the following steps of the process of the invention are described in FR-A-2 793 262.

The classification of the identical nucleic sequences in homogeneous groups is advantageously carried out by analysis of the profiles obtained by the chromatograms resulting from the DHPLC. Identical nucleic sequences are classified into homogeneous groups of DHPLC chromatograms.

Chromatography, especially DHPLC combined with sequencing, makes it possible to locate each SNP on each nucleotide fragment and to characterize the nature of the bases associated with each polymorphism.

The identification of the polymorphism of the nucleic sequence of heterozygous individuals in each group presenting a heterozygous chromatogram by comparison with the reference wild-type sequence is preferably carried out by sequencing the heterozygous nucleic sequences. Sequencing is a process well known to the one skilled in the art and here it can be carried out, for example, by the technology of capillary sequencing well known to the one skilled in the art.

By comparison with a wild-type sequence of the reference "candidate" gene, the identification of the impact of the mutant allele of each functional SNP of the nucleic sequence of each heterozygous group on the structure of the protein encoded by the "candidate" gene can be carried out by bioinformatic molecular modeling.

The present invention also has as an object a process for determination of the frequency of the polymorphism of the nucleic sequence obtained according to the above-described determination process by comparison with the reference wild-type sequence, in which one also proceeds with the genotyping of the nucleic sequences of each individual from each group of the random population obtained as explained previously.

The functional SNPs identified in the "candidate" genes in the

random population are genotyped in the same random population and a statistical analysis of the frequency of each allele (allele frequency) coded by these SNPs in the random population is then done, which makes it possible to determine the importance of their impact in the various ethnic groups that form the random population.

The genotyping data are analyzed to estimate the frequency of distributions of the different alleles observed in the populations studied. Even if the effort focusses principally on the SNPs validated functionally, the search for linkage disequilibrium between the SNPs discovered in the random population can be carried out to identify the nonfunctional SNPs that can nevertheless be associated with more relevant functional SNPs, and therefore can be markers of the latter. These nonfunctional SNPs could be used for the development of diagnostic/prognostic kits as markers of the functional SNPs with which they will be in linkage equilibrium. The calculation of the allele frequencies can be carried out with the aid of softwares such as SAS-suite® (SAS) or SPLUS® (MathSoft). The comparison of the SNPs allelic distributions through different ethnic groups of the random population can be carried out using the softwares ARLEQUIN® and SAS-suite®.

The present invention also has as an object a process for determination of the frequency of the polymorphism of the nucleic sequence identified above, in which the genotyping is carried out by minisequencing with hot ddNTPs (2 different ddNTPs labeled with different fluorophores) and cold ddNTPs (2 unlabeled ddNTPs), in combination with a polarized fluorescence reader. The minisequencing method using a polarized fluorescence reader (FP-TDI Technologie or Fluorescence Polarization Template-direct Dye-Terminator Incorporation) is well known to the one skilled in the art.

It is carried out on a product obtained after PCR amplification of the DNA of each individual, this PCR product being chosen to cover the gene region containing the SNP studied as indicated in Figure 1. After the last step of the PCR in the thermocycler, the plate is then placed on a polarized fluorescence reader for reading the labeled bases by using the excitation and

emission filters specific for the fluorophores. The intensity values of the labeled bases are reported on a graph. Thus, up to four categories are obtained, as indicated in Figure 3.

The sense and antisense primers used in the case of the human interferon α 2 gene correspond to the sequences ID SEQ No. 5 and ID SEQ No. 6, respectively.

The present invention also has as an object the use of the process for determination of the polymorphism in the nucleic sequence of a "candidate" gene described previously for the search of a variation in the nucleic sequence of a "candidate" gene. By "variation" is understood a modification of the nucleic sequence of a "candidate" gene as, for example, the presence of one or more SNP polymorphisms. The present invention therefore also has as an object the genetic diagnosis of a disease linked to the presence of the mutant allele coded by the functional SNP in one or more individuals of the human population.

The present invention also makes it possible to carry out a genetic diagnosis of a disease linked to the presence of one or several mutation(s) in the form of one or several mutant allele(s) coded by one or several functional SNP(s), to form a map of functional genetic markers taken in reference, as well as to reveal a transgenic sequence (i.e. different from the reference sequence) carried by said mutant allele in the nucleic sequence of a "candidate" gene.

The present invention also makes it possible to form a map of functional genetic markers taken in reference for the development of pharmacogenetic tests, or in other words pharmacogenomic tests, for which genetic profiling of the individuals recruited for clinical trials will be carried out from the functional SNP markers taken in reference in order to identify the panel(s) of markers that will make it possible to differentiate the responding individuals, the nonresponders or the individuals in whom the therapeutic molecules tested will have adverse effects, with the goal of optimizing said clinical trials for better effectiveness of the therapeutic molecules.

The present invention also makes it possible to develop therapeutic molecules such as antibodies, vectors for gene therapy and active

molecules determined from the structure of the mutated protein(s) encoded by the mutated allele(s) coded by one or more mutation(s) of the functional SNP type related with the appearance of or resistance to common diseases in the population, for the treatment of these same diseases.

Just as much, the present invention has as an object the use of the above process for determination of the functional SNP in the nucleic sequence of a "candidate" gene for revealing the functional SNPs in the sequence carried by said "candidate" gene existing in a random population. This also makes it possible to predict the impact of the identification of functional SNP for the diagnosis/prognosis or the treatment of these different ethnic groups.

Just as much, the present invention has as an object the use of the above process of determination of functional SNPs in the nucleic sequences of "candidate" genes for revealing or determination of new potential diagnostic/prognostic or therapeutic targets in a random population for the prevention and treatment of common diseases.

Likewise, the present invention has as an object a process for determination of the functionality of a mutant protein derived from the nucleic sequence determined by the process described above, in which the functionality of the protein derived from said nucleic sequence is compared with the functionality of the reference wild-type protein derived from the reference wild-type nucleic sequence of the "candidate" gene.

The present invention also has as an object the use of the above process for determination of functional SNP in the nucleic sequence of a "candidate" gene for the determination of the functionality of said mutated genetic sequence coded by the mutant allele coded by the functional SNP by comparing the functionality of the protein derived from said mutated nucleic sequence with the functionality of the protein derived from the reference wild-type nucleic sequence of the "candidate" gene. The determination of the functionality of a nucleic sequence depends on the nucleic sequence taken as reference and called "candidate" gene. Tools, for example bioinformatic tools,

enable a selection of the functional SNPs that are located in the regulatory sequences of the "candidate" genes which reveal a change in sequences known from the prior art as being important for the gene expression including, without being restricted to it, the TATA and CAT boxes and sites known as "enhancers".

A selection is also made of the functional SNPs that are located in the coding sequences of the "candidate" genes and that reveal the appearance of a STOP codon in these sequences and therefore an abnormal stop of the translation at the site of the functional SNP(s). Finally, among all the identified SNPs, a selection is made between on the one hand, the coding SNPs that cause a change in the nature of the amino acids of the proteins encoded by these genes and, on the other hand, the SNPs that do not cause a change in the nature of the amino acids of the proteins encoded by these genes.

The nature of the change in the sequence makes it possible to determine whether or not there is a coding of a different amino acid, and if it is different, one can examine whether this amino acid is essential to the function fulfilled by the corresponding protein.

Indeed, the physicochemical nature of the changes in the amino acids revealed by the coding SNPs can be determined, including the appearance or change in electric charge of the amino acid and the change of the hydrophilic or hydrophobic nature of the amino acid. The important amino acids and/or the domains, for which a relationship with a functional activity of the protein has been proven or is suspected, are identified. Practically, that consists of listing all the proteins belonging to the same family in the human species or in the animal kingdom and, therefore, sharing the same functional activities (homologous, heterologous or orthologous) and often a similar structure, at least for one or more domains, then generating multiple alignments. In addition, several databases are available in the public domain which list these functional domains in the form of motifs, patterns or matrices (PROSITE, BLOCKS, PFAM, etc.). An exhaustive search in the literature completes the whole information and particular attention is focused on works

disclosing observed or site-directed mutagenesis induced mutations and their involvement in the reported function of the protein. Functional SNPs found in the sequence of these important amino acids are particularly studied.

From the "candidate" gene sequence, it is possible to determine the genomic organization of the gene to be studied, to localize the promoters, the exons and the introns as well as the sites known as "splicing" sites. Only the parts of the gene for which a SNP search is relevant for the partner (example: exons) are considered.

New functional SNPs are also selected among the coding SNPs when the change in the nature of the amino acid observed for a given coding SNP concerns an amino acid in the signal peptide of the protein encoded by the "candidate" gene, in the case where a signal peptide exists, making it possible to predict a change in addressing the corresponding protein or, when the coding SNP reveals the change in an amino acid which is described in the prior art as important for the structure of the corresponding protein(s).

By identifying the residues and/or domains conserved between species and/or between these proteins and/or domains, the mutations caused by the SNPs that are able to affect the functional activity of the target can thus be predicted *in silico*.

The impact of the mutant allele revealed by this last type of SNP on the functional structure of the corresponding protein is then determined, for example using a computer software allowing molecular modeling of both types of proteins encoded by the functional SNP, the reference wild-type and the mutant proteins. Here, each protein results from the expression of each of the "candidate" gene allele coded by the functional SNP.

Previous knowledge according to the prior art of the three-dimensional structure of the reference wild-type protein and, within the protein, the amino acids involved in its activity is advantageous for allowing a reliable determination of the change caused by the mutated allele coded by the functional SNP on the structure and, therefore, the function of the protein.

Also, the protein corresponding to the reference wild-type

sequence and the mutated or mutant protein corresponding to the mutant allele can be produced by known methods.

By carrying out an appropriate in vitro test, for example a biological or pharmacological test, it can be deduced if the change caused by the mutated allele of the gene modifies or does not modify howsoever the function of the protein encoded by the "candidate" gene. In vitro expression tests can also be developed (for example, tests of expression of reporter genes such as the one coding for luciferase placed under the control of the mutated regulatory sequences) aiming to identify the mutant alleles that, in the regulatory sequences of the "candidate" genes, modify the expression of said genes.

Combined with the annotations of the protein primary sequences, the structural models of the targets can be constructed by using tools for de novo computer modeling (for example: SEQFOLD/MSI), for homology (example: MODELER/MSI), for minimization of the force fields (examples: DISCOVER, DELPHI/MSI) and/or for molecular dynamics (example: CFF/MSI).

The three-dimensional structures of the variants can then be modeled and the consequences of these structural changes on the functional activity of the target predicted.

In the case of human interferon α 2 the determination of the functionality is performed, for example, by the test of the antiproliferative activity of human interferon α 2 on human tumoral Daudi cell line of the Burkitt's lymphoma (JBC Papers in Press, published on September 12, 2000 as Manuscript M006854200).

Likewise, the present invention has as an object a process for determination of the functionality of a mutant protein such as obtained by the process described above for the development of tests for the diagnostic or prognostic of common diseases.

Likewise, the present invention has as an object a process for determination of the functionality of a mutant protein such as obtained by the process described above for the development of therapeutic molecules for the

treatment of common diseases.

Another special object of the invention is the use of the process for determination of the functionality of a protein derived from the nucleic sequence obtained as defined above for the genetic diagnosis of a disease related with the presence of one or more SNP mutations.

The execution of the present invention allows the easy selection of interesting nucleic acid fragments. That is why the present invention also has as an object nucleic acid fragments, characterized in that they contain a nucleic sequence revealed by the process for determination of a variation in the nucleic sequence of a "candidate" gene defined above and especially a nucleic acid fragment containing at least the 567 base pairs of the ID SEQ No. 4 nucleic sequence of interferon α 2, in which the nucleotide A is mutated into the nucleotide G in position 211.

The nucleic acid fragments containing a nucleic sequence revealed by the process for determination of a variation in the nucleic sequence of a "candidate" gene defined above can be obtained from the reference wild-type sequence of the "candidate" gene by mutation of the base pair(s) of the SNP(s) determined above by methods well known to the one skilled in the art and in particular by site-directed mutagenesis. The nucleic acid fragment containing at least the 567 base pairs of the ID SEQ No. 4 nucleic sequence of interferon α 2, in which the nucleotide A is mutated into nucleotide G in position 211, has been obtained by changing nucleotide A into nucleotide G in this position by site-directed mutagenesis [of] the reference wild-type sequence of the "candidate" gene.

The present invention also has as an object the use of the genetic information contained in the nucleic acid fragment described above for the genetic diagnosis of diseases such as the various types of cancers, the infection by Hepatitis B and C viruses and the AIDS virus.

These nucleic acid fragments can be incorporated into vectors. That is why the present invention also has as an object a recombinant vector comprising a nucleic sequence as described above and comprising, in addition,

regulatory regions that are positioned in such a manner to allow the expression of said nucleic sequence. Different types of recombinant vectors can be used such as expression vectors in bacteria, in mammalian cells or in insect cells such as, for example, Drosophila cells.

These recombinant vectors can be used for transfecting cells so as to obtain transformed cells. Thus, the present invention also has as an object a cell line transformed with the aid of a vector as described above. Different types of cell lines can be used such as those described above.

The present invention also has as an object a protein derived from the mutated nucleic sequence obtained by the process for determination of the functional SNP in the wild-type nucleic sequence of a reference "candidate" gene described above and especially the protein corresponding to the ID SEQ No. 7 peptide sequence, in which histidine (H) is changed to arginine (R) in position 57 of the immature protein or in position 34 of the mature protein in the case of human interferon α 2.

Numerous ways exist to produce the protein described above. Preferentially, the present invention has as an object a process for the production of such a protein, in which a transformed cell line defined above is cultivated and said protein isolated from the culture medium. Such a process is well known to the one skilled in the art.

The present invention also has as an object an antibody characterized in that it is obtained by immunization of an animal with such a protein. Such a process is well known to the one skilled in the art.

The identification of these functional SNPs thus enables human genome post-genomic or post-sequencing research for the identification of new therapeutic targets, which will make it possible to develop diagnostic or prognostic kits for these diseases, as well as new therapeutic molecules.

The present invention also has as an object an active molecule characterized in that it is developed from a protein as described above for the prevention or the treatment of diseases such as the various types of cancers, the infection by Hepatitis B and C viruses, and the AIDS virus.

The present invention also has as an object a protein such as described above, used in a diagnostic or therapeutic purpose for the prevention or treatment of diseases such as the various types of cancers, the infection by Hepatitis B and C viruses, and the AIDS virus.

The present invention also has as an object host cells comprising the recombinant vector mentioned above. The introduction of the nucleic sequences described above can be carried out by methods well known to the one skilled in the art and in the laboratory manuals such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New-York (1989). The host cells can be bacteria, fungi, yeasts, insect cells, plant cells or animal cells such as CHO, COS, HeLa, C127, 3T3, BHK and HEK 293.

The proteins determined above can be used in processes to determine new compounds with a positive (activating) or negative (inhibiting) effect on the activity of said protein. Such processes involve the use of the host cells described above in the presence of candidate compounds for experimentation. The determination of the effect produced by these candidate compounds can be carried out by experimentations such as, for example, a test of binding between the candidate compound and the host cell, or a test demonstrating the activation or inhibition of a signal in the host cell for which the protein described above is responsible.

The present invention, therefore, also has as an object a method for identification of agents activating or inhibiting the above protein, comprising:
a) placing the host cells in the presence of a compound to be tested, and
b) determination of the activating effect generated by the compound to be tested on said protein.

The present invention also has as an object an activating or inhibiting agent identified by the method described above.

The present invention also has as an object a medication containing, as active ingredient, a protein defined previously.

The present invention also has as an object the use of a protein obtained by the above process for the production of a medication for the prevention or treatment of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

The preferred conditions for carrying out the process for determination of a variation in the nucleic sequence of a preselected functional "candidate" gene described above apply equally to other objects of the invention quoted above.

Figure 1 represents the principle of minisequencing that is carried out during genotyping. The nucleotides ddATP surrounded with dotted lines are labeled with the fluorophore R110*. The nucleotides ddGTP surrounded by unbroken lines are labeled with the fluorophore Tamra*.

Figure 2 represents a wild-type profile corresponding to a homozygous individual (top) and a profile corresponding to a heterozygous individual (bottom). The abscissas represent the retention time in minutes. The ordinates represent the intensity in millivolt.

Figure 3 represents the result of genotyping the interferon α 2 H57R SNP. Base 211 a→g is genotyped in antisense t→c on the GEA 008F02 PCR fragment. The ordinates represent the mP values and correspond to the R110* filter (ddTTP). The abscissas represent the mP values and correspond to the Tamra* filter (ddCTP).

- Group 1 (top left) of 232 individuals represents the TT individuals
- Group 2 (right) represents the 4 CT individuals
- Group 3 (bottom left) represents the 7 blanks
- Group 4 (middle left) represents the 3 non genotyped individuals

The following example illustrates the present invention.

Example: Determination of a variation in the nucleic sequence of the gene encoding for human interferon alpha 2 (INFα2)

Stage a): Preselection of the "candidate" gene reference sequence

The sequence and genomic organization of the gene coding for

human interferon alpha-2 have been deposited since 1994 under the name of "interferon alpha-a" in the GenBank bank of NCBI (http://www.ncbi.nlm.nih.gov/), under the code "J00207." This sequence is used as "reference wild-type sequence" and the numbering of the nucleotide positions mentioned below are related to this sequence. The coding region (CDS) of this gene comprises 567 base pairs (bp) and codes for a protein with 189 amino acids.

The alpha interferons compose an excessively close family in terms of protein sequences in man as in all higher mammals. This is completely obvious when the sequences of these proteins are aligned by a tool such as ClustalW. The H34 residue is described by J Piehler et al. (Journal of Biological Chemistry; JBC, Sept. 2000) as participating in the domain of binding of this interferon to its receptor (receptor-2 of the interferons). It is necessary to note that this same histidine in position 34 (H34) in the mature protein is in position 57 (H57) in the immature protein. Both positions could be mentioned to refer to the same histidine amino acid that is changed in the sequence of the human interferon α 2 by the functional SNP described here. The work of J Piehler consisted of carrying out systematic site-directed mutagenesis by replacing several residues of this region by alanines. In the case of the H34A mutation, J Piehler observes a significant decrease in the ability of this interferon to interact with its receptor. The structure of monomeric interferon α 2 determined by NMR is known and available in the PDB database (http://www.rcsb.org.pdb/) under the code 1ITF.

Stage b): Isolation of the genomic DNA of the functional "candidate" gene in a random population of individuals

To discover the SNPs according to the process detailed below, a population of individuals taken at random (not selected on a particular phenotypic criterion such as collection of medical, clinical, epidemiological, physiological or biological data) has been screened and called random population.

The genomic DNAs of the individuals of the tested population have been provided by the Coriell Institute in the United States.

	DESCRIPTION	NOMBRE D'INDIVIDUS
1	Pacific Islander	7
2	Iberian	10
3	Italian	10
4	Mexican	10
5	Caribbean	10
6	African-American	50
7	Caucasian	50
8	Chinese	10
9	Indo-Pakistani	9
10	Middle-Eastern	20
11	South-American (Andes)	10
12	South-American	10
13	South Asian	10
14	South West American Indians	5
15	Greek	8
16	Japanese	10

The primers used for the polymerization chain reaction (PCR) are the following ones: G008 22F and G008 22R.

The primers used to clone the gene coding for the human interferon alpha-2 are the following ones:

GenFragm	TM	start/stop	length	sequence
G008 22F	56.03	470	20	CACCCATTTCAACCAGTCTA
G008 22R	55.77	1124	19	AGCTGGCATACGAATCAAT

Notes:

F: Sense (forward)
R: Antisense (reverse)
Start/stop: beginning (sense) or stop (antisense) of the primers by comparison with the reference sequence.
Length: size of the primers

The specificity of these primers has been tested and it appeared that no other fragment of comparable size was expected other than the fragment sought. These primers allowed amplification of the fragment F22G0088GF2 (ID SEQ No. 4 with 655 bp) whose sequence is given below (in bold, the coding sequence corresponding to ID SEQ No. 3):

F22G0088GF2
cacccatttcaaccagtctagcagcatctgcaacatctaca**atggccttgacctttgctttactggtggccct**
cctggtgctcagctgcaagtcaagctgctctgtgggctgtgatctgcctcaaacccacagcctgggta
gcaggaggaccttgatgctcctggcacagatgaggagaatctctcttttctcctgcttgaaggacaga
catgactttggatttccccaggaggagtttggcaaccagttccaaaaggctgaaaccatccctgtcctc
catgagatgatccagcagatcttcaatctcttcagcacaaaggactcatctgctgcttgggatgagacc
ctcctagacaaattctacactgaactctaccagcagctgaatgacctggaagcctgtgtgatacaggg
ggtgggggtgacagagactcccctgatgaaggaggactccattctggctgtgaggaaatacttccaa
agaatcactctctatctgaaagagaagaaatacagcccttgtgcctgggaggttgtcagagcagaaat
catgagatctttttctttgtcaacaaacttgcaagaaagtttaagaagtaaggaatgaaaactggttcaac
atggaaatgattttcattgattcgtatgccagct

In the case of interferon alpha 2, two fragments have been selected and named F1 (ID SEQ No. 4) and F2 (ID SEQ No. 3). F2 (ID SEQ No. 3) covers the coding sequences of the gene. We are presenting here the results obtained during analysis of the coding fragment F2 (GEA008F02).

Materials:
Autoclaved water
10x PCR buffer (delivered with the enzyme) GIBCO
$MgSO_4$ 50 mM

Platinum Taq enzyme 5 U/µL

dNTP 100mM

F and R primers

Genomic DNA 1 ng/µL

96-well plate (Costar)

384-well plate (ABGene)

PCR reaction: x 96-well or 384-well plates per fragment to be amplified according to the number of individuals to be tested.

Product	Supplier	Reference	Used concentration	Final concentration	Vol per well (µl)
Buffer	Gibco	11304-029	10X	1X	2.5
$MgSO_4$	Gibco 50 mM	11304-029	50 mM	0.02 M	1.075
dNTP	Gibco	10297-018	10 mM	0.2 mM	0.5
Primer F	Gibco		10 µM	0.2 µM	0.5
Primer R	Gibco		10 µM	0.2 µM	0.5
H_20					14.85
Enzyme	Gibco 5U/µl	11304-029	5 U/µl	0.375 U	0.075
DNA			1 ng/µl		5
Final volume					25

Programming the thermocyclers (Tetrad MJ research):

1 cycle:	94°C	1 min
35 cycles:	94°C	15 sec
	56°C	30 sec
	68°C	1 min

After testing the PCR products on 2% agarose gel, the amplified products are denatured on Thermocyclers (Tetrad from MJ Research)

according to the cycle program:

1 cycle:	95°C	3 min
1 cycle:	95°C	1 min

followed by a series of cycles where the temperature decreases of 1.6°C/cycle till reaching 25°C.

Once denatured, the samples are multiplexed by three on 96-well plate.

Stage c: Study of the DNA sequence of each individual

The PCR products were analyzed by DHPLC (denaturing high performance liquid chromatography).

Buffer A: for 1 liter

- 250 µL acetonitrile (ACN)
- 50 mL triethylammonium (TEAA) 2 M

Buffer B: for 1 liter

- 250 mL acetonitrile (ACN)
- 50 mL triethylammonium (TEAA) 2 M

The column is equilibrated under the following buffer conditions:

- 50% buffer A
- 50% buffer B

with a program flow of 0.9 mL/min.

The performances of the column are tested:

- on the one hand, at 50ºC by injection of 5 µL pUC 18 digested by the Hae III restriction enzyme with a buffer flow of 0.75 mL/µL and a gradient of 43% buffer B and 57% buffer A,
- on the other hand, at 56 ºC by injection of 5 µL of a standard of mutation with a buffer flow of 0.9 mL/µL and a gradient of 47% buffer B and 53% buffer A.

First, the study of sequences by the software Wave Maker® (Transgénomique Inc.) gave information on the temperature and the buffer gradient according to which the samples have to be treated. Trial tests were carried out in order to establish the effective conditions for sequences analysis.

Therefore, with the temperature(s) and buffer A and B gradient conditions, 3 µL of each of the 96 samples are analyzed for 14 h in the DHPLC apparatus called Waves® (Transgénomique Inc.)

The analysis of the fragments requires specific temperatures, obtained by the software Wave Maker® (Transgénomique Inc.), accompanied by the buffer gradients indicated in the table below,

Time (min)	%A (0,025% ACN)	%B (25% ACN)	%C (75% ACN)	Flow (ml/min)
0	45	55	0	0.9
0.1	40	60	0	0.9
4.1	32	68	0	0.9
4.2	0	100	0	0.9
4.7	0	100	0	0.9
4.8	45	55	0	0.9
6.8	45	55	0	0.9

The equilibrated column is tested with conditions proposed by the Wave Maker® (Transgénomique Inc.). These conditions are made effective during the final analysis of the F2 fragment of the samples.

The chromatograms obtained are then analyzed.

The analysis of the chromatographic profiles obtained allowed the detection of the heterozygotes and the homozygotes among the individuals of the tested population on the basis of the chromatograms or also "profiles" of distinct forms. Certain profiles allowed establishment of families (groups) of individuals presenting similar chromatograms.

- A wild-type profile corresponding to a homozygous individual (chromatogram in Figure 2 (top part))

- A different profile corresponding to a heterozygous individual (chromatogram in Figure 2 (bottom part)).

Stage d: Sequencing of the DNA from each group

Next, one proceeds with capillary sequencing of the PCR products corresponding to the heterozygous profiles using the ABI-PRISM 3700 DNA sequencers.

Sequencing protocole on the basis of a 96-well plate

Purification of the PCR products:

Weigh 50 g of Biogel P100 Fine. Suspend in 1 liter of ultrapure water. Leave standing for 8 h. Shake. Fill a multiscreen "filtering bottom" plate (Biogel P100 Fine): 400 mL per well. Superimpose on a recovery plate. Centrifuge: 500 g, 3 min. Replace the recovery plate with a new Greiner plate, superimpose with the aid of a Millipore adaptor. Deposit the PCR products on the P100. Centrifuge at 500 g, 4 min. Store at -20 °C.

Sequencing reaction:

Sequencing consists of a new PCR reaction. A sequencing reaction corresponds to the following proportions: per well containing the multiplex of fragments amplified for the detection of SNP by DHPLC from three different individuals.

- 1 µL Big Dye Terminator
- 1 µL 5X buffer (tris-HCl 400 mM//MgCl2 10 mM)
- 10 ng PCR products for 100 bp (base pairs)
- 6 pmol primer
- H_2O qsp 10 µL

Centrifuge briefly.

Reaction cycles:

- Denaturation 95 ºC / 5 min
- 95 ºC / 10 sec
- Tm / 5 sec

- 60 ºC / 4 min

25 cycles. Duration: 2.5 h

Purification of the sequencing products:

Weigh 50 g of Sephadex G50 Super-Fine. Suspend in 1 liter of ultrapure water. Leave standing 8 h. Shake. Fill multiscreen "filtering bottom" plate (Biogel P100 Fine): 400 mL per well. Superimpose on a recovery plate. Centrifuge: 1500 g, 2 min. Replace the recovery plate with a new "Optical" plate special for ABI-PRISM 3700 DNA capillary sequencing machine. Add 10 µL ultra-pure water per well to the plate after the sequencing reaction. Pour the so-diluted sequencing products on the G50. Centrifuge at 1200 g, 3 min. Store at -20 °C.

Migration of the samples:

Migration is carried out on the ABI-PRISM 3700 DNA capillary sequencer.

Analyze with the following conditions: The "Optical" plate containing the samples is recovered and it is covered with an adhesive aluminum foil. Place the plate on a rack adapted for the ABI-PRISM 3700 DNA capillary sequencer and place the whole in a free carrier A, B, C or D. Verify the levels of buffer, water, polymer, isopropanol. Adjust them if necessary.

In the START menu, PE Biosystems tab, under subdirectory "3700 Programs", open "Data Collection". In the "Plate set up" tab, import the operation sheet by clicking on "import." Assign the operation sheet by clicking on the carrier containing a large question mark, carrier that corresponds to the plate to be sequenced. When it is active, click on the green arrow. Time of trial: 4 h.

Control of the sequences:

In the START menu, PE Biosystems tab, open "Data Extraxtor." Click on "Extract Now." In the START menu, BE Biosystems tab, open

"Sequencing Analysis 3.6". Click on "add files" and import the previously extracted sequences. Open the sequences one by one and verify the quality of the electrophorograms i.e. the quality of migration of the sequences in the capillaries, the reading length, estimate the percentage of readable sequences. Transfer the sequences into the computer network, file "Sequencing – Sequences Discovery", for identification of the SNPs.

With the aid of the sequences and with the "PolyPhred" software for sequences analysis, the nature of the nucleotide and the position of the polymorphism have been identified. In position 680 of the reference wild-type sequence of the gene coding for interferon alpha 2, base A is replaced by G in a pool of 3 individuals in a random population. The overlay of the peaks is informative of the SNP.

Stage e): Genotyping of a functional SNP

Once the SNP is identified, it is analyzed to identify if it changes an amino acid present in the mature protein. Amino acid change: H57R (histidine changed into arginine in position 57 of the immature protein or 34 of the mature protein).

- Technique used: fluorescent minisequencing. FP-TDI Technologie or Fluorescence Polarization Template-direct Dye-terminator Inc.
- Principle of minisequencing: SNPs genotyping is based on the principle of minisequencing for which the product is detected by polarized fluorescence reading. Minisequencing consists of elongating an oligonucleotide, placed just upstream of the polymorphic site, by fluorolabeled dideoxynucleotides with the aid of a polymerase enzyme as illustrated in Figure 1. The result of this elongation is directly analyzed by polarized fluorescence reading.

Steps of the protocol:

Minisequencing is carried out on a product obtained after PCR amplification of an interferon α 2 gene sequence fragment which carries the functional SNP from the genomic DNA from each individual of the random

population. This PCR product is chosen to cover the gene region containing the SNP studied. Then, the PCR primers and the unincorporated dNTPs are eliminated before carrying out the minisequencing. All these steps, as well as the reading, are carried out in the same plate.

Therefore, genotyping requires 5 steps:

1) Amplification by PCR
2) Purification of the PCR product by enzymatic digestion
3) Elongation of the oligonucleotide
4) Reading
5) Interpretation of the reading

1) The PCR amplification of the interferon α 2 gene sequence which covers the gene region containing the functional SNP is carried out using the same primers as those used for the identification of the SNPs. Therefore, the PCR product is made for each individual of the random population as described above in the step for the discovery of the functional SNP. This PCR product is used as template for the minisequencing reaction which is used to genotype the individuals for the functional SNP. The amplification by PCR is carried out in the same plate. The reaction volume is 5 μL as described in the following table:

Supplier	Reference	Reagent	Initial conc.	Vol. per tube (µl)	Final Conc.
Life Technologie	delivered with Taq	Buffer (X)	10	0.5	1
Life Technologie	Delivered withTaq	$MgSO_4$ (mM)	50	0.2	2
AP Biotech	27-2035-03	Dntp (mM)	10	0.1	0.2
Life Technologie	on request	F primer (µM)	10	0.1	0.2
Life Technologie	on request	R primer (µM)	10	0.1	0.2
Life Technologie	11304-029	platinum Taq	5U/µl	0.02	0.1 U/ reaction
		H_2O	Qsp 5 µl	1.98	
		DNA	2.5 ng/µl	2	5 ng/ reaction
		Final Volume		5 µl	

These reagents are distributed in a black PCR plate with 384 wells provided by ABGene (ref: TF-0384-k). Once filled, the plate is sealed, centrifuged then placed in a thermocycler for 384-well plate (Tetrad from MJ Research) and incubated in the following conditions: PCR cycles: 1 min at 94 ºC, followed by 36 cycles composed of 3 steps (15 sec at 94 ºC, 30 sec at 56 ºC, 1 min at 68 ºC).

2) The PCR is then purified using two enzymes, shrimp alkaline phosphatase (or Shrimp Alkaline Phosphatase SAP) and exonuclease I (Exo I)). The first of these enzymes allows the dephosphorylation of the dNTPs that have not been incorporated during the PCR, while the second enzyme eliminates the single-stranded DNA residues and therefore the primers that have not been used during the PCR. This digestion is carried out by adding 5 µL of reaction mixture, prepared as described in the table that follows, to the PCR plate:

Supplier	Reference	Reagent	Initial Conc.	Vol. per tube (µl)	Final Conc.
AP Biotech	E70092X	SAP	1 U/µl	0.5	0.5/ reaction
AP Biotech	070073Z	Exo I	10 U/µl	0.1	1/ reaction
AP Biotech	Delivered with SAP	SAP Buffer (X)	10	0.5	1
		H_2O	Qsp 5 µl	3.9	
		PCR		5 µl	
		Final Volume		10 µl	

Once filled, the plate is sealed, centrifuged then placed in a thermocycler for 384-well plate (Tetrad from MJ Research) and is incubated in the following conditions: SAP-EXO digestion: 45 min at 37 ºC, 15 min at 80 ºC.

3) The elongation or minisequencing step is then carried out on this digested PCR product by the addition of a reaction mixture prepared as given in the table below:

Supplier	Reference	Reagent	Initial Concentration	Vol. per tube (µl)	Final Concentration
own preparation		Elongation buffer* (X)	5	1	1
Life Technologies	On request	Miniseq primer (µM)	10	0.5	1
AP Biotech	27-2051 (61, 71,81)-01	**ddNTPs (µM) (2 cold ddNTPs)	2.5 of each	025	0.125 of each
NEN	Nel 472/5 and Nel 492/5	**ddNTPs (µM) (2 labeled ddNTPs Tamra and R110)	2.5 of each	0.25	0.125 of each
AP Biotech	E79000Z	Thermo-sequenase	3.2 U/µl	0.125	0.4 U/ reaction
		H2O	Qsp 5 µl	3.125	
		Digested PCR		10 µl	
		Final volume		15 µl	

* The 5X elongation buffer is composed of 250 mM Tris-HCl pH 9, 250 mM KCl, 25 mM NaCl, 10 mM $MgCl_2$ and 40% glycerol

** For the ddNTPs, a mixture of the 4 bases is carried out according to the polymorphism studied. Only the 2 bases of interest (A/G) composing the functional SNP bear a labeling either with Tamra or R110 ex SNP A/G: the mixture of ddNTPs is composed of:

- 2.5 µM cold ddCTP,
- 2.5 µM cold ddTTP,
- 2.5 µM ddATP (1.825 µM cold ddATP and 0.625 µM Tamra-labeled ddATP),
- 2.5 µM ddGTP (1.825 µM cold ddATP and 0.625 µM R110-labeled ddATP).

Once filled, the plate is sealed, centrifuged, then placed in a thermocycler for 384-well plate (Tetrad from MJ Research) and incubated in the following conditions: Elongation cycles: 1 min at 93 ºC, followed by 35 cycles composed of 2 steps (10 sec at 93 ºC, 30 sec at 55 ºC).

After the last step in the thermocycler the plates is directly placed on an Analyst® HT polarized fluorescence reader from LJL Biosystems Inc. The plate is read with the aid of the Criterion Host® software by using two methods. The first method allows the reading of the Tamra-labeled base using excitation and emission filters specific for this fluorophore (excitation 550-10 nm, emission 580-10 nm) and the second method allows the reading of the R110-labeled base using the excitation and emission filters specific for this fluorophore

(excitation 490-10 nm, emission 520-10 nm). In both cases, a dichroic double mirror (R110/Tamra) is used and the other reading parameters are:

Z-height: 1.5 mm

Attenuator: out

Integration time: 100,000 µsec

Raw data units: counts/sec

Switch polarization: by well

Plate settling time: 0 msec

PMT setup: Smart Read (+), sensitivity 2

Dynamic polarizer: emission

Static polarizer: S

A result file is then obtained, that contains the mP values calculated for the Tamra filter and those for the R110 filter. These mP values are calculated from values of intensity obtained on the parallel plane (//) and on the perpendicular plane (⊥) according to the following formula:

$$mP = 1000(// - g.\perp)/(// + g.\perp).$$

In this calculation the value on the filter ⊥ is weighted with a factor g. This is a parameter specific of the apparatus, that must be previously determined experimentally.

4) and 5) Interpretation of the reading and determination of the genotypes

The mP values are reported on a graph using the Excel software from Microsoft Inc., or now using the Allele Caller® software developed by LJL Biosystems Inc. On the abscissa is given the mP value of the Tamra-labeled base, on the ordinate is given the mP value of the R110-labeled base. A high mP value indicates the incorporation of the base labeled with this fluorophore and, conversely, a low mP value reveals the absence of incorporation of this base. Up to four categories are obtained, as given in Figure 1. Once the different groups are determined, the use of the Allele Caller® software allows to directly extract the genotype

defined for each individual in the form of a table.

The sequences of the two minisequencing primers necessary for the genotyping have been determined. These primers are selected to correspond to the 20 nucleotides placed just upstream of the polymorphic site. Because the PCR product containing a SNP is a double-stranded DNA product, the genotyping can therefore be carried out either on the sense strand or the antisense strand. The primers selected are produced by Life Technologies Inc. The minisequencing of the SNP A211G on the GEA008F02 fragment was first validated on 16 samples then genotyped on the entire random population composed of 239 individuals and 10 blanks.

The minisequencing primers are as follows:

Sense primer (ID SEQ No. 5) GEA008F02A211UP: ctcctgcttgaaggacagac

Antisense primer (ID SEQ No. 6) GEA008F02A211LO: cctggggaaatccaaagtca

Minisequencing conditions tested:

Condition No. 1: sense primer + ddATP-R110 + ddGTP-Tamra

Condition No. 2: sense primer + ddGTP-R110 + ddATP-Tamra

Condition No. 3: antisense primer + ddTTP-R110 + ddCTP-Tamra

Condition No. 4: sense primer + ddCTP-R110 + ddTTP-Tamra

These four conditions have been tested and condition No. 3 has been retained for genotyping.

Results:

Genotyping of the random population was carried out using the condition described previously. The genomic DNA of the individuals of the random population (see stage b) of Example 1) were provided by the Coriell Institute of the United States.

After completion of the genotyping process, the determination of the genotypes of the individuals of the random population for the functional SNP studied here was carried out using the graph represented in Figure 3. This

genotype is in theory either homozygous AA, or heterozygous AG, or homozygous GG in the individuals tested. In reality and as shown below, the homozygous GG genotype is not detected in the random population.

The results for the controls, the distribution of the genotypes determined in the random population and the calculation of the different allele frequencies for this functional SNP are presented in the following table:

Number of individuals		**Number of blanks**		**Percentage of success**
tested	genotyped	tested	validated	
239	236	7	7	99.2

Distribution of genotypes		
Number of TT	Number of TC	Number of CC
232 (on the left of the graph)	4 (on the right of the graph)	0

Genotype Frequency (%)			**Allele frequency (%)**	
TT	TC	CC	T	C
98.3	1.7	0	99.2	0.8

Definition of the allele or genotype frequency: it is the frequency of a given allele or genotype estimated in a population.

It is necessary to specify that allele T read in antisense corresponds to allele A read in sense, that is to say to the presence of a histidine in position 57 of the INF alpha 2 and therefore that the allele C read in antisense corresponds to the allele G read in sense corresponding to an arginine for this position in the corresponding protein sequence.

By examining these results by population it is noted that the 4 heterozygous individuals all come from a single subpopulation or ethnic group, the "African American" subpopulation of the random population. The analysis of this functional SNP in this population is as follows:

Distribution of genotypes			Genotype frequency (%)			Allele frequency (%)	
Number of TT	Number of TC	Number of CC	TT	TC	CC	T	C
45	4	0	91.8	82	0	95.9	4.1

Claims

1. Process for determination of one or more functional SNP polymorphisms in the nucleic sequence of a preselected "candidate" gene in which:

a) the genomic nucleic acid fragment of the "candidate" gene is isolated from a significant number of individuals chosen randomly in the population,

b) a comparative analysis of the nucleic sequence of the individuals studied is conducted,

c) the identical nucleic sequences are classified into homogeneous groups, and

d) the functional SNP of the nucleic sequence of the heterozygous group(s) is identified by comparison with the nucleic sequence of the reference "candidate" gene.

2. Process for determination according to Claim 1, in which the "candidate" gene is preselected by carrying out a search in the literature or in the databases.

3. Process for determination according to Claim 1 or 2, in which the "candidate" gene may be involved in the appearance of or resistance to a particular pathology.

4. Process for determination according to any one of Claims 1 to 3, in which the "candidate" gene is the human interferon α 2 gene.

5. Process for determination according to any one of Claims 1 to 4, in which the significant number of individuals chosen randomly in the population is greater than 100.

6. Process for determination according to any one of Claims 1 to 5, in which the "candidate" gene's nucleic sequence of a significant number of individuals chosen randomly in the population is isolated by a PCR reaction.

7. Process for determination according to Claim 6, characterized in that the PCR is carried out from primers corresponding to the sequences ID SEQ No. 1 and ID SEQ No. 2.

8. Process for determination according to any one of Claims 1 to 7, in which the comparative analysis of the nucleic sequence of the individuals studied is carried out by a multiplexing method using denaturing high performance liquid chromatography (DHPLC).

9. Process for determination according to any one of Claims 1 to 8, in which the classification of the identical nucleic sequences in homogeneous groups of homozygous and heterozygous is carried out by analysing the profiles obtained from the DHPLC chromatograms.

10. Process for determination according to any one of Claims 1 to 9, in which the identification of the two alleles of each functional SNP of the nucleic sequence of each heterozygous group by comparison with a wild-type sequence of the reference "candidate" gene is carried out by sequencing the nucleic sequences or fragments of nucleic sequences.

11. Process for determination according to any one of Claims 1 to 10, in which the identification of the impact, on the structure of the protein encoded by the "candidate" gene, of the mutant allele of each functional SNP of the nucleic sequence of each heterozygous group by comparison with a wild-type sequence of the reference "candidate" gene is carried out by bioinformatic molecular modeling.

12. Process for determination of the frequency of the polymorphism of the nucleic sequence obtained according to the process for determination according to any one of Claims 1 to 11 by comparison with a wild-type sequence of the reference "candidate" gene, in which, in addition, one proceeds with genotyping of the individuals of a random population for the alleles of the functional SNP obtained according to the process for determination according to any one of Claims 1 to 11.

13. Process for determination in the random population of the allele and genotype frequency of the functional SNP of the nucleic sequence according to Claim 12, in which the genotyping is carried out by minisequencing.

14. Process for determination of the allele and genotype frequency of the functional SNP of the nucleic sequence according to Claim 13,

characterized in that the sense and antisense primers used correspond to the sequences ID SEQ No. 5 and ID SEQ No. 6, respectively.

15. Use of the process for determination of the functional SNP in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for searching a sequence variation in a "candidate" gene.

16. Use of the process for determination of the functional SNP in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for the genetic diagnosis of a disease related to the presence of the mutant allele coded by the functional SNP in one or more individuals of the human population.

17. Use of the process for determination of the functional SNP in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for making a map of genetic markers.

18. Use of the process for determination of a variation of functional SNP-type in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for revealing a transgenic sequence carried by said "candidate" gene.

19. Use of the process for determination of a variation of functional SNP-type in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for revealing all the sequence polymorphisms of functional SNP-type carried by said "candidate" gene in a given population.

20. Process for determination of the functionality of a protein derived from the sequence of mutant allele coded by a functional SNP determined according to any one of Claims 1 to 14, in which the functionality of the protein derived from said nucleic sequence is compared with the functionality of the protein derived from the reference wild-type nucleic sequence of the "candidate" gene.

21. Use of the process for determination of the functional SNP in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14 for the determination of the functionality of said genetic sequence coded by the mutated allele by comparing the functionality of the protein derived from said

mutated nucleic sequence with the functionality of the protein derived from the reference wild-type nucleic sequence of the "candidate" gene.

22. Use of the process for determination of the functionality of a protein derived from the nucleic sequence obtained according to Claim 20 for the genetic diagnosis of a disease related to the presence of one or more mutation(s) of functional SNP-type.

23. Use of the process for determination of the functionality of a protein derived from the nucleic sequence obtained according to Claim 20 for the development of therapeutic molecule such as an antibody, a vector for gene therapy, and an active molecule determined from the structure of the mutated protein(s) encoded by the mutated allele(s) coded by one or more mutation(s) of functional SNP-type.

24. Nucleic acid fragment, characterized in that it contains at least the nucleic sequence revealed by the process for determination of a variation of functional SNP-type in the nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14.

25. Nucleic acid fragment containing at least the 567 base pairs of the nucleic sequence ID SEQ No. 4, in which nucleotide A is mutated into nucleotide G in position 211.

26. Use of the genetic information contained in the nucleic acid fragment of Claim 24 or 25, for the genetic diagnosis of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

27. Recombinant vector comprising a nucleic sequence according to Claim 24 or 25 and additionally comprising regulatory regions that are positioned in such a manner that the expression of said nucleic sequence is possible in bacteria, in mammalian cells, or insect cells.

28. Cell line transformed by the recombinant vector according to Claim 27.

29. Protein derived from the nucleic sequence according to Claim 22 or according to the process for determination of the functional SNP in the

nucleic sequence of a "candidate" gene according to any one of Claims 1 to 14.

30. Protein according to the peptide sequence ID SEQ No. 7, in which histidine (H) is changed to arginine (R) in position 57 of the immature interferon α 2 protein or in position 34 of the mature interferon α 2 protein.

31. Process for production of the protein defined in Claim 29 or 30, in which a cell line according to Claim 28 is cultivated and said protein is isolated from the culture medium.

32. Antibody, characterized in that it is obtained by immunization of an animal with a protein defined in Claim 29 or 30.

33. Antibody, characterized in that it is obtained by immunization of an animal with a protein defined in Claim 29 or 30, with a diagnostic or therapeutic purpose for the prevention or treatment of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

34. Active molecule, characterized in that it is developed from a protein defined in Claim 29 or 30 for the prevention or treatment of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

35. Protein defined in Claim 29 or 30, used within a diagnostic or therapeutic purpose for the prevention or treatment of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

36. Host cells comprising the recombinant vector according to Claim 27.

37. Method for identification of agents activating or inhibiting the protein defined in Claim 29 or 30, comprising:

a) placing the host cells according to Claim 36 in the presence of a compound to be tested, and

b) the determination of the activating or inhibiting effect generated by the compound to be tested on said protein.

38. Activating or inhibiting agent identified by the method

according to Claim 37.

39. Medication containing a protein defined in Claim 29 or 30 as active ingredient.

40. Use of a protein obtained according to Claim 29 or 30, for the production of a medication for the prevention or treatment of diseases such as the various types of cancers, the infection by the Hepatitis B and C viruses, and the AIDS virus.

SEQUENCE LISTING

<110> GENODYSSEE S.A.

<120> Process for determination of one or more functional polymorphisms in the nucleic sequence of a preselected functional « candidate » gene, and its applications.

<130> genodyssee

<140>
<141>

<160> 7

<170> PatentIn Ver. 2.1

<210> 1
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic oligonucleotide

<400> 1
cacccatttc aaccagtcta 20

<210> 2
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic oligonucleotide

<400> 2
agctggcata cgaatcaat 19

<210> 3
<211> 567
<212> DNA
<213> Homo sapiens

<400> 3
atggccttga cctttgcttt actggtggcc ctcctggtgc tcagctgcaa gtcaagctgc 60
tctgtgggct gtgatctgcc tcaaacccac agcctgggta gcaggaggac cttgatgctc 120
ctggcacaga tgaggagaat ctctcttttc tcctgcttga aggacagaca tgactttgga 180
tttccccagg aggagtttgg gaaccagttc caaaaggctg aaaccatccc tgtcctccat 240
gagatgatcc agcagatctt caatctcttc agcacaaagg actcatctgc tgcttgggat 300
gagaccctcc tagacaaatt ctacactgaa ctctaccagc agctgaatga cctggaagcc 360
tgtgtgatac agggggtggg ggtgacagag actcccctga tgaaggagga ctccattctg 420
gctgtgagga aatacttcca aagaatcact ctctatctga aagagaagaa atacagccct 480

tgtgcctggg aggttgtcag agcagaaatc atgagatctt tttctttgtc aacaaacttg 540
caagaaagtt taagaagtaa ggaatga 567

<210> 4
<211> 655
<212> DNA
<213> Homo sapiens

<400> 4
cacccatttc aaccagtcta gcagcatctg caacatctac aatggccttg acctttgctt 60
tactggtggc cctcctggtg ctcagctgca agtcaagctg ctctgtgggc tgtgatctgc 120
ctcaaaccca cagcctgggt agcaggagga ccttgatgct cctggcacag atgaggagaa 180
tctctctttt ctcctgcttg aaggacagac atgactttgg atttccccag gaggagtttg 240
ggaaccagtt ccaaaaggct gaaaccatcc ctgtcctcca tgagatgatc cagcagatct 300
tcaatctctt cagcacaaag gactcatctg ctgcttggga tgagaccctc ctagacaaat 360
tctacactga actctaccag cagctgaatg acctggaagc ctgtgtgata caggggtgg 420
gggtgacaga gactcccctg atgaaggagg actccattct ggctgtgagg aaatacttcc 480
aaagaatcac tctctatctg aaagagaaga aatacagccc ttgtgcctgg gaggttgtca 540
gagcagaaat catgagatct ttttctttgt caacaaactt gcaagaaagt ttaagaagta 600
aggaatgaaa actggttcaa catggaaatg attttcattg attcgtatgc cagct 655

<210> 5
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Description of the artificial sequence: synthetic oligonucleotide

<400> 5
ctcctgcttg aaggacagac 20

<210> 6
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Description of the artificial sequence: synthetic oligonucleotide

<400> 6
cctggggaaa tccaaagtca 20

<210> 7
<211> 188
<212> PRT
<213> Homo sapiens

<400> 7
Met Ala Leu Thr Phe Ala Leu Leu Val Ala Leu Leu Val Leu Ser Cys
1 5 10 15

Lys Ser Ser Cys Ser Val Gly Cys Asp Leu Pro Gln Thr His Ser Leu
20 25 30

Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Lys Ile Ser
35 40 45

Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu
50 55 60

Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu His
65 70 75 80

Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser
85 90 95

Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr Thr Glu Leu Tyr
100 105 110

Gln Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly Val
115 120 125

Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys
130 135 140

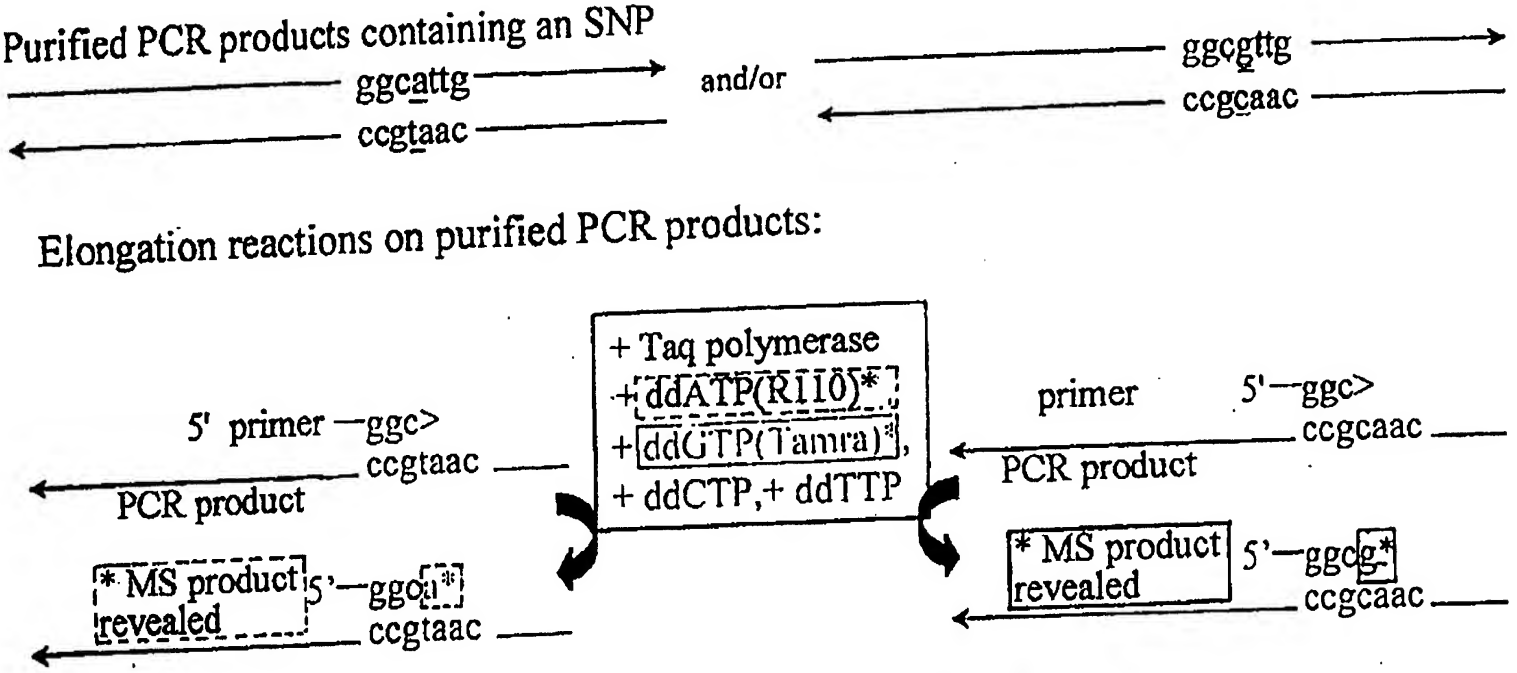
Tyr Phe Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro
145 150 155 160

Cys Ala Trp Glu Val Val Arg Ala Glu Ile Met Arg Ser Phe Ser Leu
165 170 175

Ser Thr Asn Leu Gln Glu Ser Leu Arg Ser Lys Glu
180 185

FIGURE 1

Minisequencing



Individual a-a: labeled with R110*

Individual a-g: labeled with R110* + Tamra*

Individual g-g: labeled with Tamra*

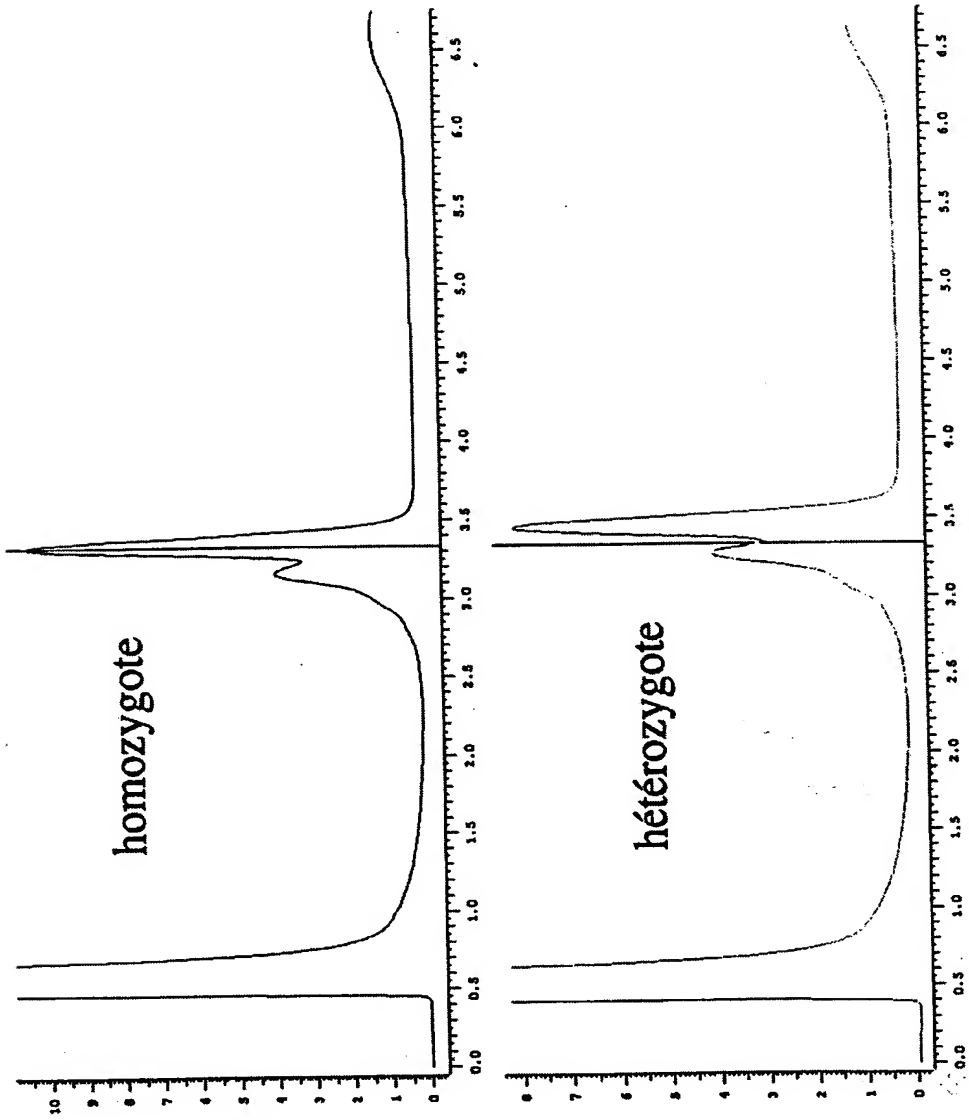


FIGURE 2

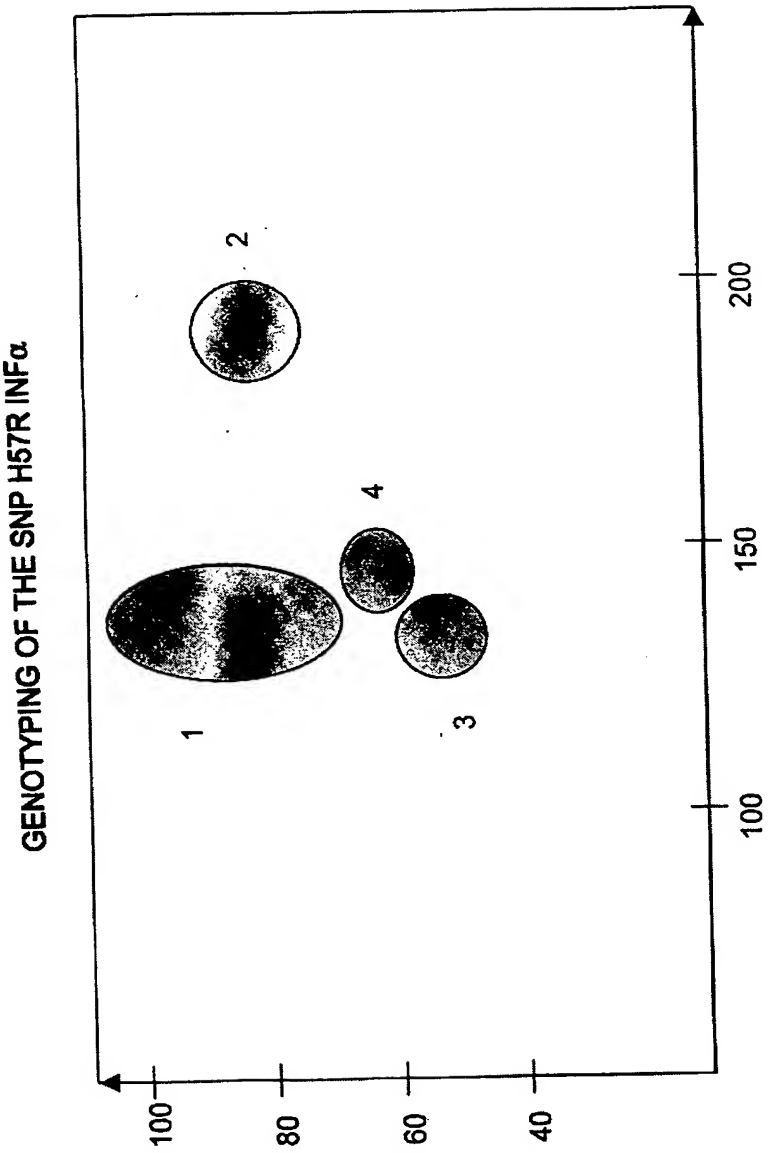


FIGURE 3

DESCRIPTIVE ABSTRACT

APPLICANT

Corporation known as: GenOdyssee

Agents

RINUY, SANTARELLI

CLAIM OF PRIORITIES

See the title of the invention and the text of the abstract enclosed

TITLE OF THE INVENTION

Process for determination of one or more functional polymorphism(s) in the nucleic sequence of a preselected functional "candidate" gene and its applications.

TEXT OF THE ABSTRACT

Process for determination of one or more polymorphisms of functional SNP-type in the nucleic sequence of a preselected "candidate" gene in which:

a) the genomic nucleic acid fragment of the "candidate" gene is isolated from a significant number of individuals chosen randomly in the population,

b) a comparative analysis of the nucleic sequence of the individuals studied is conducted,

c) the identical nucleic sequences are classified into homogeneous groups, and

d) the functional SNP of the nucleic sequence of the heterozygous group(s) is identified by comparison with the nucleic sequence of the reference "candidate" gene.